Inverse Quantitative Structure Activity Relationship Analysis for Improving Predictions of Chemical Toxicity

Shawn Martin
Sandia National Laboratories
(505) 284-3601
smartin@sandia.gov

Authors: Shawn Martin[1], John Kenneke[2], W. Michael Brown[1], Jean-Loup Faulon[3]
[1]Sandia National Laboratories, Albuquerque, NM
[2]U.S. EPA National Exposure Research Laboratory, Ecosystem Research Division, Athens, GA
[3]Sandia National Laboratories, Livermore, CA
Keywords: toxicity, prediction, quantitative structure activity relationships, metabolism, computation

The toxic outcomes associated with environmental contaminants are often not due to the chemical form that was originally introduced into the environment, but rather to the chemical having undergone a transformation prior to reaching the vulnerable species. More importantly, the chemical is often transformed (or metabolized) to the toxic form inside the species of interest. This situation is so common that any tool for accurately predicting toxicity must include a module that accurately predicts metabolism. In response to this need, the U.S. National Exposure Research Laboratory (NERL) Ecosystem Research Division (ERD) in Athens, GA is developing a metabolic simulator in support of the U.S. Environmental Protection Agency (U.S. EPA) Office of Research and Development (ORD) Computational Toxicology Program.

In a joint project between the U.S. EPA and Sandia National Laboratories, inverse quantitative structure activity relationships (QSARs) are being used to elucidate structural motifs that lead to activated, and potentially harmful, metabolites. Utilization of these QSARs in the design and development of the metabolic simulator will result in greater accuracy than current chemoinformatic methods, provide a source for a universal descriptor from which other descriptors could be computed, provide a means to control descriptor degeneracy, and be used to generate molecular structures (i.e., new chemicals), which will then be used to test the metabolic simulator and target areas requiring further research or data.

We are in the process of developing the inverse QSAR method. As a proof of concept, we have used 27 conazole fungicides and corresponding fish Chronic Toxicity Values (ChV) computed using the U.S. EPA's Persistent, Bioaccumulative and Toxic (PBT) profiler. We have trained a QSAR using forward-stepping multilinear regression and signature. The final QSAR used six signature descriptors and achieved a Q2 value from leave-one-out cross validation of 0.65 and an R2 value of 0.91, both indicating predictive capability. To invert the QSAR, we derived 38 linear Diophantine equations relating the total 107 signature descriptors. Although we were unable to enumerate completely the solutions to these equations using our standard method (a Contejean–Devie solver), we nevertheless obtained 10,047 solutions using an alternative method under development. Our next step will be reconstructing full structures as well as sub-motifs from these 10,047 solutions as described above. These structures and motifs will be evaluated using the QSAR and then used to test the metabolic simulator.

Although this work was reviewed by the U.S. Environmental Protection Agency and approved for publication, it may not necessarily reflect official Agency policy.